Abstract:
Reducing the energy consumption of datacenters and the Cloud is very important in order to lower CO_2 footprint and operational cost of a Cloud operator. However, there is a tradeoff between energy consumption and perceived application performance. In order to save energy, Cloud operators need to consolidate as many Virtual Machines (VM) on the fewest possible physical servers. However, such consolidation policy may lead to violation of SLA if VMs suddenly run at their peak demand and the number of used servers was too small to cope with such overload. Such consolidation is facilitated in modern datacenters by using live VM migration techniques, which stress the network. As a consequence, it is important to find the right balance between the energy consumption of the physical servers, the placement of the VMs after the consolidation process and the number of migrations to perform. What makes such VM consolidation challenging is the service reliability and SLA requirements of VMs given the highly dynamic nature in modern datacenters, where many aspects have a high degree of uncertainty. For example, the resources that a VM requires are not precisely known in advance, they may change during the day depending on workload. Another uncertainty factor is the energy consumption model of the rack servers which makes it difficult to precisely predict the energy demands due to many optimizations in modern hardware. Unfortunately, the presence of uncertain data in an optimization problem may lead to solutions that are useless in practice. This is because small deviations in input data values may lead to situations where a found optimal solution is even not feasible any more.

In this work, we therefore propose a novel approach to the energy efficient VM consolidation problem based on the theory of Robust Optimization (RO). The key idea of RO is that input to an optimization model (e.g. parameters of the objective function, constraint matrix, or right hand side) is not known precisely, but rather support functions are used to define a scenario where such parameters vary within given bounds. Given the fact that for large number of unknown variables it is very unlikely that the whole uncertainty is taking place at the same time over all coefficients one can effectively solve such RO model. By applying this theory, we develop a mathematical model of the energy efficient VM consolidation problem as a robust Mixed Integer Linear Program and specify uncertainty bounds for resource requirements of the VMs and the rack server power model. Those bounds are inspired by a measurement campaign done in our universities datacenter observing CPU workload of 6 different VMs that implement various parts of our economic system and authentication infrastructure. We show that our model allows to tradeoff two important aspects for a datacenter operator: by taking higher risk aversion, our model will take into account more severe and unlikely deviations of the uncertain parameters, leading to higher protection but also higher energy consumption. Alternatively, if one wants to take a higher risk, the solution will offer less protection at lower energy consumption.